D1-82-0345

BOEING SCIENTIFIC RESEARCH LABORATORIES

BLANK PAGE

Still Another Method for Producing Normal Variables in a Computer

DDC
MAR 8 1965
DDC-IRA B

George Marsaglia

Mathematics Research

January 1965

D1-82-0345

STILL ANOTHER METHOD FOR PRODUCING

NORMAL VARIABLES IN A COMPUTER

by

George Marsaglia

SUMMARY

A method for producing normal random variables in terms of uniform random variables U_1, U_2, U_3, $\cdots$. If $Y = U_1 + U_2 + U_3$, then choosing one of the four random variables $2Y - 3$, $(4Y - 6)/3$, $(Y - 7)/2$ or $(Y + 4)/2$ in the proportions .8635, .11506, .00372 and .00372 will produce the required normal variate 98.6 per cent of the time. The other 1.4 per cent is devoted to the tail or a rejection technique in order that the composite be exact. The method leads to very fast computer programs which are easy to code and occupy little space in the computer.

THE METHOD

The following method for generating normal random variables in a computer is similar to that reported in reference [2]. It is slightly faster and simpler. Programs based on this method take about half again as much time as the very fastest methods, [1]. The fastest methods, however, require a few thousand storage locations and are relatively complicated. The basic normal random variable generator for a large computer probably should use the fastest available method, but in situations where ease of programming or storage requirements are more important, the following method is particularly well suited.

We want to generate a normal random variable X, density $(2\pi)^{-.5}e^{-.5x^2}$, by using independent random variables U_1, U_2, U_3,..., uniformly distributed over the interval $(0,1)$. We assume our computer has a subroutine for providing a satisfactory sequence of U's. The idea is to use the sum of 3 uniform variables, $Y = U_1 + U_2 + U_3$, in one of the four linear forms $2Y - 3$, $(4Y - 6)/3$, $(Y - 7)/2$, and $(Y + 4)/2$. These particular functions of Y are chosen because they are easy to generate and it is possible to approximate the normal density very closely with a mixture of their densities.

If $f(y)$ is the density of $Y = U_1 + U_2 + U_3$,

$$f(y) = \begin{cases} .5y^2 & 0 \leq y \leq 1 \\ .5y^2 - 1.5(y - 1)^2 & 1 \leq y \leq 2 \\ .5y^2 - 1.5(y - 1)^2 + 1.5(y - 2)^2 & 2 \leq y \leq 3 \\ 0 & y < 0 \text{ or } y > 3, \end{cases}$$

then the densities of $2Y - 3$, $(4Y - 6)/3$, $(Y - 7)/2$, and $(Y + 4)/2$ are $.5f(.5X + 1.5)$, $.75f(.75x + 1.5)$, $2f(2x + 7)$, and $2f(2x - 4)$. We use the following mixture of these densities to approximate the normal density:

$$(1) \qquad (2\pi)^{-.5}e^{-.5x^2} \cong .8635[.5f(.5x + 1.5)] + .11506[.75f(.75x + 1.5)] +$$
$$.00372[2f(2x - 4)] + .00372[2f(2x + 7)].$$

Furthermore, the expression on the right is always less than $(2\pi)^{-.5}e^{-.5x^2}$; the difference is plotted in Figure 1. This means that we can generate X from one of the four densities on the right of (1) some 98.6 per cent of the time, and the remaining 1.4 per cent of the time we generate X from either the tail or the residual portion of the density. The combined mixture provides an X with density exactly $(2\pi)^{-.5}e^{-.5x^2}$. The method for handling X from the tail of the normal distribution is a slight modification of that described in [3], and the rejection technique is used to provide X from the residual density on $-3.5 \leq x \leq 3.5$. See Figure 1. Our method for generating X is then to choose one of six alternate ways--from one of the four densities on the right of (1), from the residual density of Figure 1, or from the tail. The details are in this outline:

To generate a standard normal random variable X, density
$\varphi(x) = (2\pi)^{-.5}e^{-.5x^2}$, let $Y = U_1 + U_2 + U_3$ and

1. With probability .8635, put $X = Y + Y - 3$

2. With probability .11506, put $X = 4Y/3 - 2$

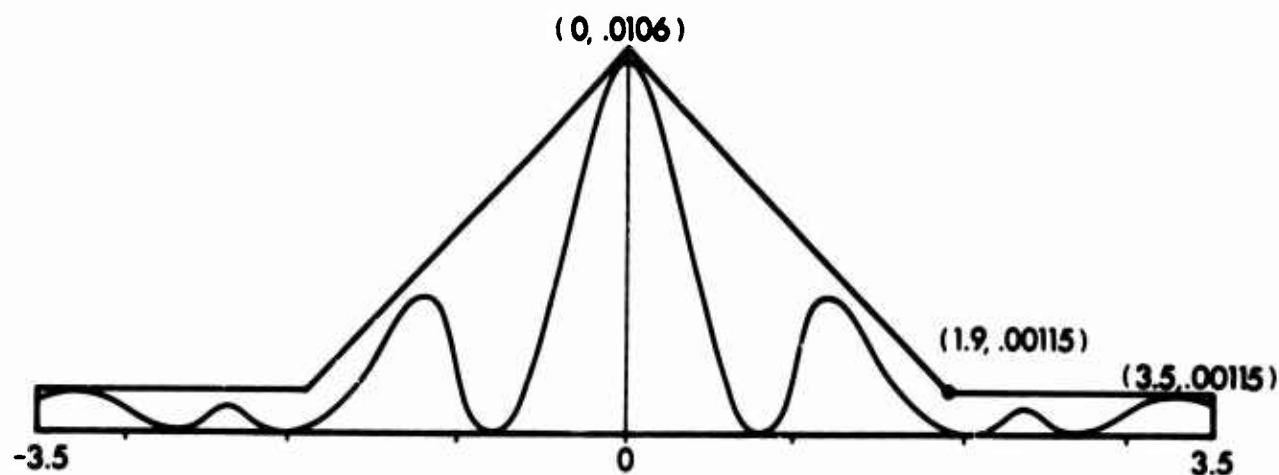3. With probability .00372, put $X = .5Y - 3.5$

Figure 1. Graph over $-3.5 < x < 3.5$ of the residual function $(2\pi)^{-.5}e^{-.5x^2} - .43175f(.5x+1.5) - .086295f(.75x+1.5) - .00744[f(2x-4)+f(2x+7)]$. The area under the curve is $.013534.7418$. To generate a variate X with density the appropriate multiple of this function, generate (x,y) uniformly from the hat-shaped region until (x,y) lies under the curve, then put $X = x$. Choose (x,y) uniformly from the hat-shaped region by putting $x = 7U_1 - 3.5$, $y = .00115U_2$, with probability $.3095558546$ or $x = 7.0(U_1 + U_2 - 1)$, $y = .00115 + .00945U_3(1 - |U_1 + U_2 - 1|)$ with probability $.6904441454$. Efficiency $= \frac{.013535}{.026005}$, or 52 per cent.

4. **With probability** .00372, **put** $X = .5Y + 2$

5. **With probability** $.014 - 2 \int_{3.5}^{\infty} \varphi(t)dt = .0135347418\cdots$ **choose** (x,y) **uniformly from the hat-shaped region of Figure 1 until**

$$y < (2\pi)^{-.5} e^{-.5x^2} - .43175f(.5x + 1.5) - .086295f(.75x + 1.5)$$
$$- .00744[f(2x - 4) + f(2x + 7)]$$

then put $X = x$.

6. **With probability** $2 \int_{3.5}^{\infty} \varphi(t)dt = .0004652582\cdots$ **generate pairs** $x = 2U_1 - 1$, $y = U_2$ **until** $y < 3.5(12.25 - 2 \ln|x|)^{-.5}$, **then put**

$$X = \begin{cases} (12.25 - 2 \ln|x|)^{.5} & \text{if } x > 0 \\ \\ -(12.25 - 2 \ln|x|)^{.5} & \text{if } x < 0 \end{cases}$$

The values in this outline are exact, so that combining the 6 steps with the indicated frequencies will produce an X with density exactly $\varphi(x)$. Practically, of course, one is restricted by the word length of the computer for the U's and the probabilities in steps 5 and 6.

The fastest method, [1], will produce normal variables in the IBM 7094 computer at the rate of 12,500 per second, while the above method will produce 8,500 per second. In the IBM 1620 decimal computer the fastest method yields 50 variates per second versus 32 for the above method.

REFERENCES

[1] G. Marsaglia, M. D. MacLaren, T. A. Bray, "A Fast Procedure for Generating Normal Random Variables," Comm. Assoc. Comp. Mach., Vol. 7 No. 1, 1964, 4-10.

[2] G. Marsaglia, T. A. Bray, "A Convenient Method for Generating Normal Variables," SIAM Review, Vol. 6 No. 3, July 1964, 260-264.

[3] G. Marsaglia, "Generating a Variable from the Tail of the Normal Distribution," Technometrics, Vol. 5 No. 1, 1964, 101-102.